DOCUMENT RESUME

ED 048 345 TM 000 418

AUTHOR Jacobs, Stanley S.
TITLE An Experimental Analysis of Answer-Changing Behavior on Objective Tests.
PUB DATE Feb 71
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Achievement Tests, Educational Experiments, *Graduate Students, *Multiple Choice Tests, *Response Style (Tests), Student Ability, *Student Behavior, *Test Wiseness

ABSTRACT

The study was an experimental investigation of the effects of item difficulty and subject ability on subjects answer-changing behaviors. Subjects were administered an achievement test composed of items at three levels of difficulty via slides, followed by a printed copy of the test. Analyses revealed no effects attributable to subject ability. Item difficulty was related to both frequency and qualty of change. Fewest answers were changed for easiest items, with the greatest number of changes and points gained on moderately difficult items. A generally inverse relationship appeared between quality of change and difficulty. Subjects were unable to predict the outcome of their answer-changing; while approximately 50% felt they typically lost, on the average all subjects gained regardless of their opinion. (Author)

ED0 48345

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

AN EXPERIMENTAL ANALYSIS OF ANSWER-CHANGING BEHAVIOR ON OBJECTIVE TESTS

Stanley S. Jacobs

University of Pittsburgh

TM 000 418

Paper presented at the annual convention of the American Educational Research Association

New York, New York

February 1971

AN EXPERIMENTAL ANALYSIS OF ANSWER-CHANGING BEHAVIOR ON OBJECTIVE TESTS

Stanley S. Jacobs

University of Pittsburgh

The study was an experimental investigation of the effects of item difficulty and subject ability on Ss answer-changing behaviors. Ss were administered an achievement test composed of items at three levels of difficulty via slides, followed by a printed copy of the test. Analyses revealed no effects attributable to subject ability. Item difficulty was related to both frequency and quality of change. Fewest answers were changed for easiest items, with the greatest number of changes and points gained on moderately difficult items. A generally inverse relationship appeared between quality of change and difficulty. Ss were unable to predict the outcome of their answer-changing; while approximately 50% felt they typically lost, on the average all Ss gained regardless of their opinion.

There seems to be some feeling on the part of students that initial decisions concerning objective test items are usually correct, although apparently the only published data on this point are those of Mathews (1929). Mathews' data, however, fail to support students' opinions, i.e., it is apparently advisable to change one's responses, since the typical result is an improvement in test score.

Writers on the topic of test-taking behaviors are not always in agreement concerning the advisability of answer-changing. Huff (1961), in

a widely-read guide for test-takers, implies that it is usually inadvisable to change answers. Millman, Bishop and Ebel (1965), however, suggest that the tendency to evaluate and judiciously change one's item responses is a basic aspect of test-wiseness.

A number of studies (Lehman, 1928; Mathews, 1929; Jarrett, 1948; Reile and Briggs, 1952; Bath, 1967) have concluded that there is a relationship between total test scores and the quality of changes made. That is, better students gain more than poorer students when answers are changed. However, one must be aware of a possible tautology since the stratifying variable (total test score) is simply the summation of item scores which are effected by the changes made.

These studies, as well as those of Berrein (1939) and Lowe and Crawford (1939) have demonstrated that the general result of answer-changing behavior is a higher test score.

The present study was designed to investigate the inter-relationship of item difficulty, ability level of Ss and answer-changing behavior on an objective achievement test, with some degree of control maintained over the decision-making process.

Methodology

Subjects

The sample of 50 Ss involved in the present study were drawn from the enrollment of the introductory graduate course in educational research at the University of Pittsburgh. Participation in research was a part of course requirements.

Procedure

In the first week of the term, all Ss completed the Quick Word Test (QWT) (Borgatta and Corsini, 1964), a 100 item, 4-option multiple-choice

vocabulary test. Following the first course examination, Ss were requested to compose a brief note detailing their opinion concerning answer-changing on objective tests, ending with a statement indicating whether S felt the net result was a gain or loss in test score, or whether the result was unknown to S.

Approximately four weeks later, Ss completed an examination dealing with measurement concepts composed of 45 4-option multiple-choice items. The items were drawn from a larger pool of items for which item analysis data were available so that the test contained 15 easy items ($\bar{p}$ = .75), 15 items of moderate difficulty ($\bar{p}$ = .49) and 15 very difficult items ($\bar{p}$ = .29). All items were positively discriminating and an attempt was made to maintain content validity. Items were randomly ordered within the test.

Items were reproduced singly on 2 x 2 slides. Items with a total word count of 25 or greater were produced as black on white slides, and given an exposure time of 45 seconds. Items with a word count of less than 25 were produced as white on black slides, and exposed for 30 seconds. Ss were informed of the mode of testing and the exposure times and cues. They were informed they would see the slides only once, to read the items rapidly but carefully, and to answer all items.

Slides were presented using a remotely-controlled Kodak Carousel Model 850 projector with an Endalight screen. Timing was done using a Sears Model 19902 stop-watch.

Upon completion of the 45 item test, Ss were informed they would have the opportunity to reconsider their answers. Black electrographic pencils used to complete the test were collected, and Ss received a mimeographed

copy of the test and a red pencil. Any changed answers were to be recorded in red, without erasing initial responses, allowing the determination of the frequency and quality of changed answers.

Analysis

A 2 x 3 x 3 three dimensional chi-square was developed with the following dimensions.

(1) ability: Ss were divided at the median of the QWT scores into low and high ability groups.

(2) type of change: Wrong-to-right, right-to-wrong, and wrong-to-wrong categories were established for changed responses.

(3) item difficulty: Items were categorized as being of low, moderate or high difficulty, based upon analysis information provided by a similar group two terms earlier.

A two-way ANOVA for repeated measures was employed to analyze net gains realized through answer-changing, as a function of subject ability and level of item difficulty. Extreme groups of n=15 were formed for the ability variable.

A one-way ANOVA was employed to analyze net gains made by Ss previously reporting gain, loss, or no decision concerning their answer-changing behavior. Due to absences when the initial reports were collected, the n for this analysis is 44 rather than 50.

Results

Of the five chi-squares calculated, only two were significant at the .05 level: the χ^2 between dimensions (2) and (3), and the total χ^2. Since the other dimensions appeared independent and the interaction χ^2 was non-sig-

nificant, the total χ^2's significance may be attributed to the dependence between dimensions (2) and (3). (See Table 1)

TABLE 1

Frequency of Types of Answer Changes

Made to Items of Low, Moderate and High Difficulty

Type of Change	Level of Item Difficulty		
	Low	Moderate	High
Right-to-wrong	41	50	57
Wrong-to-right	134	184	93
Wrong-to-wrong	20	58	98

$\chi^2 = 68.2$, $p < .05$

As may be seen in Table 1, there is a marked tendency to change incorrect responses to correct responses, with the quality of changes showing a gradual deterioration as item difficulty increases. As one might expect, fewest answers are changed for the easiest items, and the amount gained is least for the difficult items. (See Table 2)

TABLE 2

Summary of Net Gains Resulting from Changes, for

Three Levels of Item Difficulty and Two Levels of Subject Ability

Level of Ability	Level of Item Difficulty					
	Low		Moderate		High	
	$\bar{x}$	s.d.	$\bar{x}$	s.d.	$\bar{x}$	s.d.
High	1.93	2.12	2.47	1.81	.33	1.91
Low	1.47	1.73	3.47	2.72	1.20	2.43

As summarized in Table 3, there is no significant difference attributable to subject ability in net gains when answers are changed, but there are significant differences attributable to the level of item difficulty. The greatest gains are realized when Ss change the answers to items of moderate difficulty, the least when answers to very difficult items are changed.

TABLE 3

Repeated Measures ANOVA Summary Table for Effects of Subject Ability and Level of Item Difficulty on Net Gain Scores[1] (Conservative Test (Winer, 1962))

Source	df	MS	F
Ability (A)	1	4.90	.89
$Error_A$	28	5.48	
Level of Diff. (B)	2	36.58	8.82*
A x B	2	4.93	1.19
$Error_B$	56	4.15	

Students are apparently unable to predict the outcome of their answer-changing behavior accurately. As seen in Tables 4 and 5, all groups gain as a result of answer-changing, and the differences among groups are non-significant.

[1]Scheffe's test, for repeated measures, showed the locus of the significant difference to be between moderately difficult and highly difficult items only.

TABLE 4

Summary of Net Gains Over Total Test for

Those Ss Reporting a Typical Gain, Loss, or No Opinion

Gain			Loss			Do Not Know		
n	$\bar{x}$	s.d.	n	$\bar{x}$	s.d.	n	$\bar{x}$	s.d.
13	6.0	2.9	20	4.8	3.1	11	5.5	4.3

TABLE 5

One-way ANOVA Testing Differences in Actual

Gains Among Groups Reporting Gain, Loss, or No Opinion

Source	df	MS	F
Between	2	6.34	0.56
Within	41	11.27	

Discussion of Results

Although the generalizability of the present study may be somewhat limited due to the unique testing procedure employed, it was deemed of greater importance to first insure some degree of internal validity for the study. The previous work cited depended upon post-hoc examinations of test papers to determine frequency and quality of answer change. Aside from the questionable reliability of the procedure, it seems to assume that if any answers are changed, the student must put pen to paper. In view of the question of "overt" versus "covert" changes of mind, it was decided to

(1) imply to students that they would see items only once, (2) pace them carefully through those items and (3) force a response to the item, followed by an opportunity to reconsider initial answers in a manner of readily detectable by the experimenter.

The present study indicates that, student opinion notwithstanding, students should be allowed and encouraged to reconsider answers to multiple-choice items. The improvement in scores may be greatest on somewhat speeded tests composed of moderately difficult items. If one were interested in the best approximation of a "true score," it would seem advisable to reduce the degree of speededness as much as possible. It also appears that verbal ability of the type measured in the present study is unrelated to gains made in answer changing. The question of achievement of Ss and gains was not investigated. However, one must be aware of a possible "ceiling effect," i.e., better students may make far fewer changes, thereby gaining less.

REFERENCES

Bath, J.A. Answer changing behavior on objective examinations. Journal of Educational Research, 1967, 61, 105-107.

Berrein, F.K. Are first impressions best on objective tests? School and Society, 1939, 50, 319-320.

Borgatta, E.F. and Corsini, R.J. Quick Word Test. New York: Harcourt, Brace & World, Inc., 1964.

Huff, D. Score -- the strategy of taking tests. New York: Ballantine Books, 1961.

Jarrett, R.F. The extra-chance nature of changes in students' responses to objective test items. Journal of General Psychology, 1948, 38, 243-250.

Lehman, H.C. Does it pay to change initial decisions in a true-false test? School and Society, 1928, 28, 456-458.

Lowe, M.S. and Crawford, C.C. First impression versus second thought in true-false tests. Journal of Educational Psychology, 1929, 20, 192-195.

Mathews, C.O. Erroneous first impressions on objective tests. Journal of Educational Psychology, 1929, 20, 280-286.

Millman, J., Bishop, C.H., and Ebel, R. An analysis of test-wiseness. Educational and Psychological Measurement, 1965, 25, 707-726.

Reile, P.J. and Briggs, L.J. Should students change their initial answers on objective-type tests?: More evidence regarding an old problem. Journal of Educational Psychology, 1952, 43, 110-115.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.